A Data-Centric Framework for Diagnosing False Data Injection Attacks in Smart Grids

Michael Kallitsis, Shrijita Bhattacharya, George Michailidis
mgkallit@merit.edu, shrijita@umich.edu, gmichail@ufl.edu

***Abstract*—The smart power grid is a "system of systems" whose two-way communication capabilities play a critical role in the grid's secure, reliable and efficient operation. At the same time, the data communication functionalities introduced to Advanced Metering Infrastructure (AMI) nodes end the grid's isolation, and expose the network into an array of cyber-security threats that jeopardize the grid's stability and availability. For instance, malware amenable to inject *false data* into the smart meter monitoring infrastructure can compromise demand response schemes and endanger the grid's state estimation process or even induce power outages. In this paper, we explore a data-centric framework, coupled with an array of *spatiotemporal models*, for efficient diagnosis of *false data injection attacks* in smart grids. We leverage the co-linearities arising in the power consumption of smart meters in close proximity (e.g., belonging to the same neighborhood-area network) and propose methods that rapidly detect suspicious behavior due to possible data attacks. We discuss the necessary data "pre-processing" steps required by the proposed state-space models, and demonstrate, *with real-world power data obtained from a large university campus*, a sequential hypothesis testing framework and visualization tools that can assist utility operators in detecting and diagnosing such cyber-attacks on their networks.**

***Index Terms*—Anomaly detection, false data injection attacks, dynamic factor models, state-space algorithms, correlative monitoring, smart grid, real-data.**

I. Introduction

Modernizing the aging electric network with advanced metering infrastructure constitutes a fundamental milestone posited by utility companies. Over the past few years, millions of "smart" meters have been deployed [1]; by 2020, an estimated 90 million devices would be installed in the U.S. alone [2]. These next-generation AMI meters are equipped with low-latency two-way communication functionalities that enable advanced meter diagnostics (e.g., real and reactive power, voltage and current, frequency, etc.), efficient and more accurate accounting and billing, rapid troubleshooting for outage remediation, and, most importantly, network controllability [3]–[5]. A prime motivation for the latter is *demand response* schemes [6]–[8], which are already in place by several utilities[1]. Reacting to sudden elevation of power demand is either impossible (generating backup power cannot usually be achieved in a timely manner) or prohibitively expensive to power companies [9]. Thus, utilities strongly favor "shedding" unnecessary load in order to maintain power consumption at desired levels. Via accurate state-estimation and load forecasting, and by employing the feedback-loop that next-generation smart meters allow, demand response mechanisms can sustain the grid's reliable and energy-efficient operation.

The integrity of reported data is, thus, critical for the grid's secure operation. Erroneous data could compromise the grid's state estimation process and this might lead to brown-outs or black-outs [5], [10]–[12]. Nefarious actors interested in disrupting the network's smooth operation could simply launch the so-termed *false data injection (FDI) attacks* [13], [14] in a coordinated manner to introduce network instabilities, with or *without* complete knowledge of the network's topology (e.g., see [15], [16]). In such attacks, one could rapidly increase or decrease the power load reported by smart meters in order to imperil the statistical methods used to infer the grid's state, convey falsified information and mislead network operators. Given the pervasive use of smart meters in today's grid, it is perhaps not surprising that adversarial remote operation of critical AMI infrastructure (and similar attacks) have already been documented [17]–[20].

In this paper, we propose a generic framework (see Fig. 1) and a series of statistical models and corresponding estimation algorithms aiming towards rapid detection of FDI attacks. Our methodology is purely data-centric—no information on network topology is required—and behavioral-based—no prior knowledge of attack signatures is necessary. We follow a *correlative*

[1]Examples include, but are not limited to, the Pacific Northwest DR project (https://www.nwcouncil.org/energy/dr) and the New England DR initiative (http://nedri.raabassociates.org).

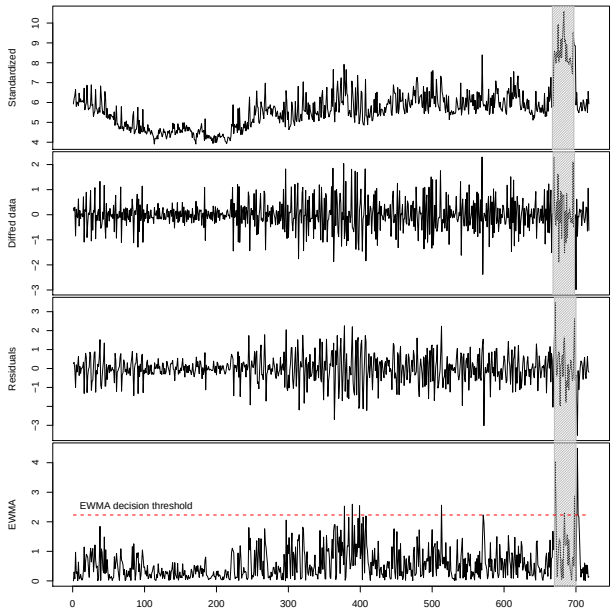

Fig. 1: The main steps of our framework. First, pre-processing steps such as *standardizing* and *differencing* are required in order to eliminate scaling and non-stationarity, respectively, issues. Then, state-space models, such as the three methods discussed in this paper, can be employed to forecast electricity consumption based on past observations. The *residuals* (errors) between predicted and meter-reported values are treated as *monitoring statistics*, which are passed into a sequential hypothesis testing framework for detection and visualizations. The figure above displays the standardized and de-trended power consumption (in Watts), the prediction errors and the smoothed monitoring statistic (via an exponentially weighted moving average) over time.

monitoring approach; we employ power consumption data from AMI meters in the same "neighborhood-area" network [21], and leverage the fact that such data are spatio-temporally correlated (see Fig. 2). We consider *univariate and multivariate autoregressive* models, as well as *dynamic factor models* [22], [23] suitable when only a few "common factors" can explain a large portion of power load variability. We also examine a *majority voting* variant, in which the ensemble of algorithms notifies for anomalies only when at least two alerts are "active". All models are integrated in a sequential hypothesis testing framework for online identification of injection attacks.

Our main *contributions* are: a) following a data-centric approach and being topology-agnostic, we study statistical models for accurate detection of data attacks. The key insight we seek to convey is that short-term power forecasting can help analysts diagnose meter anomalies by checking for significant deviations between predicted and meter-reported values (see Fig. 1); b) we evaluate all models using *real-world electricity consumption data*; to the best of our knowledge, no prior work has examined such data in the context of FDI attacks; c) we provide a generic framework, recommendations and lessons-learned to utility operators for handling non-stationary, high-dimensionality data with seasonalities for the task at hand.

Note that the proposed framework offers a principled *workflow* for addressing the FDI problem. For example, one may employ other variants of spatio-temporal models; e.g., Bayesian versions of the proposed models [24], nonlinear state space models [25] etc. The particular choice of the models presented in this study is due to their conceptual simplicity, good overall performance and ease to estimate their parameters fast. Analogously, variants of the proposed testing framework may be used. Finally, the novel idea of using ensemble methods for the problem at hand offers robust performance and is broadly applicable.

II. Related Work

Methodologies for anomaly detection in the electric grid can in general be categorized into signature-based, specification-based and anomaly-based techniques [21], [26], [27]. The proposed approach falls within the latter category. A known disadvantage of signature-based methods is that new attacks with patterns agnostic to signature-based intrusion detection systems (e.g., *Snort*) would always evade detection. At the same time, specification-based systems [28] can be cumbersome to fine-tune (e.g., finding a valid range for the AMI demand / supply is not easily determined). On the other hand, behavioral-based (i.e., anomaly-based) methods like ours can adopt to available data sources and construct a "data subspace" that describes normal operations; when data significantly deviate from this subspace, alerts for abnormal behavior are raised.

Data attacks have been studied significantly over the past few years [29]–[39]. In [29], the problem of detecting aberrant behavior of residential smart meters is tackled from the *home-area network* perspective. In [30], the authors employ a "network kriging" model to detect attacks on AMI meters based on observations from a subset of "trusted" nodes. [31] presents a mixed-integer programming method that determines the smallest subset of measurements that need to be protected to render FDI attacks ineffective. [32] proposes an adaptive cumulative sum test combined with a multivariate hypothesis testing

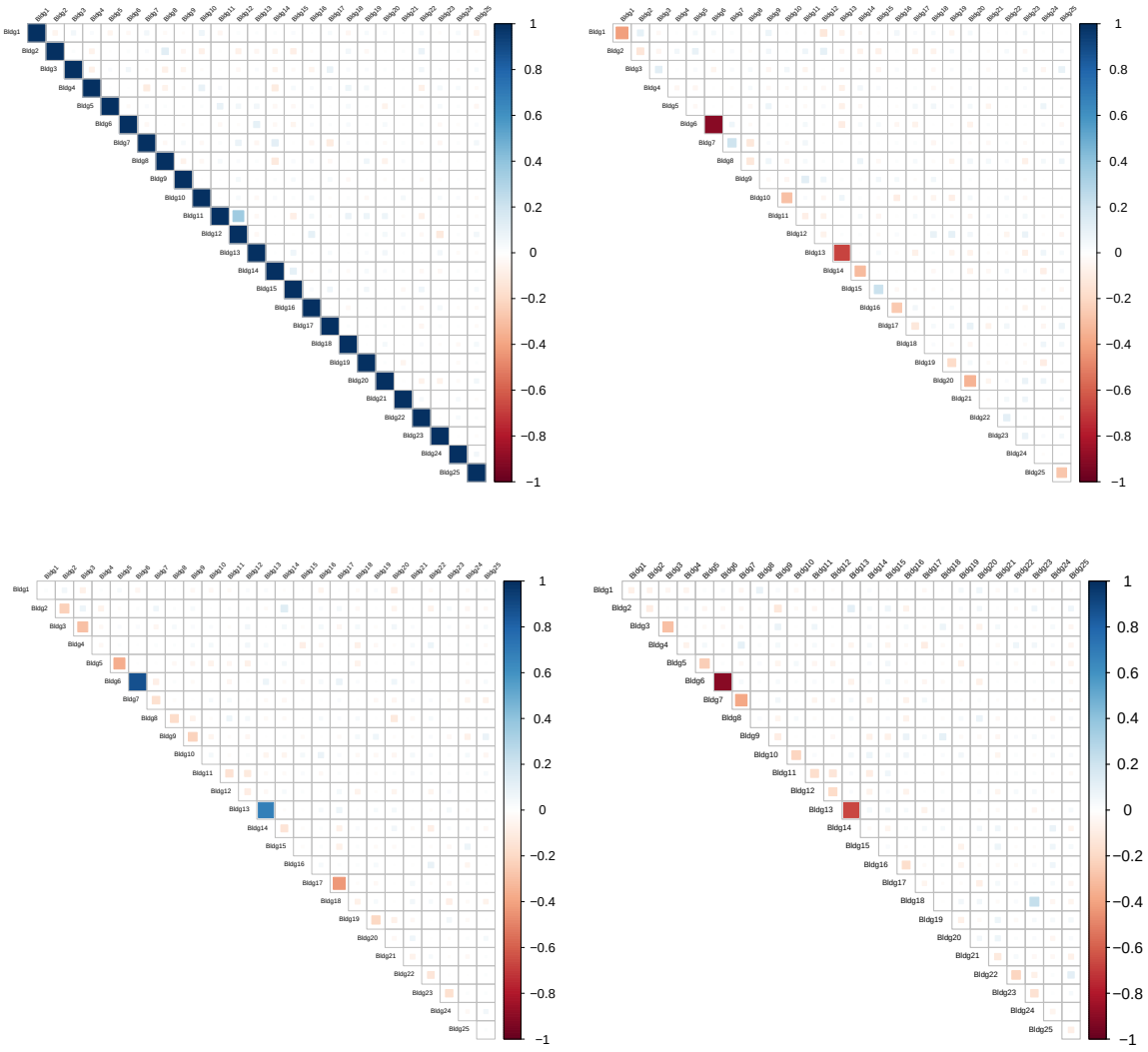

Fig. 2: Spatiotemporal correlation among the buildings (de-trended signal, 25 buildings displayed). Correlation plots for the first 3 lags (lag 0, 1 shown in first row; lags 2, 3 in second).

problem to prevent an erroneous grid-state estimate. [33] studies a graph theoretic method for securing an optimal set of meter measurements so that state estimation is not compromised. [34] couples anomaly-based methods with a data integrity check to combat stealth attacks. [35] sheds light into situations of *multiple adversaries* performing injection attacks, and discusses optimal defense strategies using a game theoretical framework. A linear measurement model (namely, a vector autoregressive one) is derived in [36] to handle both SCADA and PMU (phasor measurement unit) measurements. In [37], detection of FDI attacks is formulated as a matrix-factorization problem. The work in [38] establishes a sequential hypothesis testing framework based on the likelihood ratio that is amenable to distributed realization. A multilayer neural-network is studied in [39], trained as a binary classifier to identify the presence of FDI attacks.

Contrary to the related work in [13], [32]–[34], [36], [40], we emphasize that our method does not require parameter knowledge of the system's *DC power flow model* (see [5], [13]) nor the grid topology; instead, the

power utilization of the AMI meters constitutes the sole input.

III. Problem Formulation and Framework

The proposed framework receives as input time-series of AMI data, such as power consumption data reported by meters to "data aquisition" databases. The goal is to provide online notifications to utility operators for possible suspicious power consumption activity, such as FDI attacks. The latter is achieved by raising an alert whenever a suitable *monitoring statistic* exceeds an analyst-controlled threshold.

The techniques presented below can be implemented in an online manner, by iterating between the following three primary steps (see also Fig. 1): a *data preprocessing* step, a *forecasting* step for obtaining an estimate of the power consumption for the next-step ahead, and a *statistical hypothesis test*. Based on a suitable modeling distribution, the hypothesis test module checks whether the error/difference between the predicted and the actual electricity consumption seems anomalous (out-of-control of the normal operating range).

The discussion below sheds more light into the proposed system. For forecasting, we propose and thoroughly explore *three* different statistical models. However, as noted in the introduction, analysts may "plug-in" alternative forecasting techniques and apply the same detection framework.

A. Dealing with Non-stationary Data

The preprocessing module employs techniques that help tackle the non-stationary nature of AMI data, inherent in power consumption data for the time span of interest (e.g., several hours or days). One can observe trends (e.g., an upward trend is evident in our data of university dormitories, as students return back from school breaks) or seasonalities (for example, periodicities due to diurnal effects—see Fig. 3). However, most statistical models heavily rely on the assumption of stationarity. Hence, to properly leverage such data, one first needs to appropriately *transform* them, and then work with the *mean-corrected* transformed series. As suggested by [41], [42], successively differencing the data series yields a new, de-trended signal that one can work with. After extensive experimentation, we concluded that first-order differences, e.g., $\nabla X_t = X_t - X_{t-1}$, are sufficient for the task at hand when working with short time-series (i.e., 1–4 days or few hours). Such a preprocessing step removes short-term upward/downward trends. To remove seasonalities, [41] suggests to difference the series at lags equal to the data period d. We experimented with such transformations, but did not observe any additional advantage for the purposes of anomaly detection. Indeed, Fig. 4 illustrates that no hidden periodicities are present after de-trending the series with just first differences. If longer-term power forecasting was our primary focus, then lag-d differencing might have proved useful. Fig. 1 also shows that the differenced signal reveals no patterns and exhibits characteristics of a stationary series. Various *goodness-of-fits* criteria employed (e.g., see Fig. 4 and Fig. 6) demonstrate that our models were appropriately trained.

Next, we discuss three time series models used for forecasting purposes. Two of them (dynamic factor and vector autoregressive models) capture different aspects of the cross-correlation structure in the multiple series data, while the third and simplest one (autoregressive model) only focuses on each series own temporal dependence characteristics.

B. Dynamic Factor Model

The first forecasting method explored for purposes of anomaly detection is the *dynamic factor model* (DFM), (see Doz *et al.* [22], [23]). DFM models are heavily employed in forecasting economic indicators; however, to the best of our knowledge, such techniques have not been explored for modeling electricity data. Our DFM algorithm is outlined below, and follows the work in [22], [23].

1) The Model: Let N be the number of AMI meters in the monitoring set, and $X_t = (X_{1t}, X_{2t}, \ldots, X_{Nt})^\top$ the corresponding N-dimensional time series observed at equally-spaced time intervals. We posit the following linear model (we follow the notation in [23]) for the power consumption data X_t, $t = 1, 2, \ldots$,

$$X_t = \Lambda_0 F_t + \xi_t, \tag{1}$$

where Λ_0 is the $N \times r$ matrix of *factor loadings*, $F_t = (f_{1t}, \ldots, f_{rt})^\top$ is a stationary process of *common factors*, and $\xi_t = (\xi_{1t}, \ldots, \xi_{Nt})^\top$ is a stationary process of *idiosyncratic* components. (F_t) and (ξ_t) are assumed independent. The observed process X_t is, thus, decomposed into two *latent* orthogonal components; a *common* component, F_t, driven by (few) $r \ll N$ factors that capture most of the correlation across time series, and an *idiosyncratic* one, ξ_t, modeled as Gaussian white noise with zero mean and covariance Ψ_{0d}. To capture the dynamics of the factors, we assume that process (F_t)

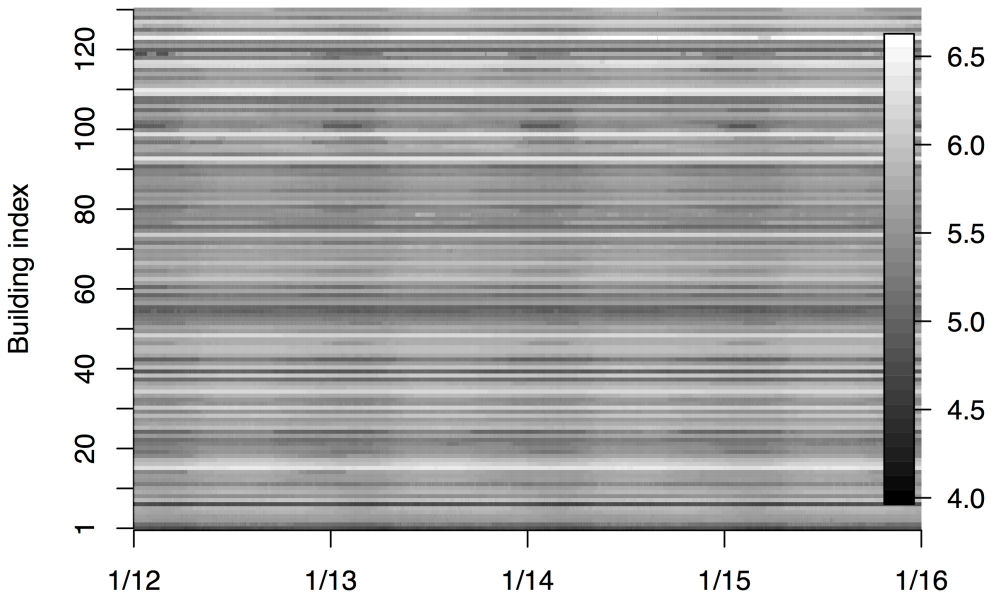

Fig. 3: Heatmap of building power consumption (Watts, log-scale). Observe the diurnal trend (periodic "white shade").

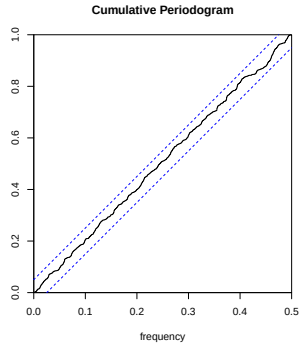

Fig. 4: The cumulative periodogram [42] of residuals verifies no hidden periodicities exist after de-trending the series.

admits a VAR(p) (vector auto-regressive model of order p) representation, namely

$$A_0(L)F_t = w_t, \tag{2}$$

where $A_0(L)$ is a matrix polynomial in L [43], L is the back-shift operator, i.e., $LF_t = F_{t-1}$, and (w_t) is a sequence of independent and identically distributed (Gaussian) random vectors with zero mean and covariance Q_0.

2) Learning the Model Parameters: The model described above is fully-specified if the parameters $\{\Lambda_0, A_0(L), \Psi_{0d}, Q_0\}$ were known[2]. In practice, though, only the observations $X_t, t = 1, \ldots$, are available. However, estimates of the aforementioned parameters can be obtained via *principal component analysis (PCA)* [44], [45]. One obtains estimates of Λ_0, using a "training" set of observations $t = 1, \ldots, T$, by considering the *empirical* covariance matrix, S, of the (centered) data. To obtain the PCA-based estimate, let $S = \frac{1}{T}\sum_{t=1}^{T} X_t X_t^\top$. We denote the *singular value decomposition* $S = PDP^\top$, where $D = \mathrm{diag}(d_1, d_2, \ldots, d_N)$ is a diagonal matrix of eigenvalues in decreasing order, and P a matrix whose columns are the eigenvectors p_j corresponding to the eigenvalues d_j, $j = 1, \ldots, N$.

Let $\hat{D} = \mathrm{diag}(d_1, d_2, \ldots, d_r)$ be the $r \times r$ diagonal matrix of the r-largest eigenvalues, and $\hat{P} = \mathrm{diag}(p_1, p_2, \ldots, p_r)$ the associated $N \times r$ matrix. Then, the PCA-based solution yields, for $t = 1, \ldots, T$,

$$\hat{F}_t = \hat{D}^{-1/2}\hat{P}^\top X_t \tag{3}$$

$$\hat{\Lambda} = \hat{P}\hat{D}^{1/2}. \tag{4}$$

From (1), and since without loss of generality $\mathbb{E}(F_t F_t^\top) = I_r$, the covariance of the idiosyncratic components is estimated as $\Psi = S - \hat{\Lambda}^\top\hat{\Lambda}$. However, in lieu of working with idiosyncratic components with cross-sectional correlation, we will assume a diagonal covariance matrix, namely we set $\Psi_d = \mathrm{diag}(S - \hat{\Lambda}^\top\hat{\Lambda})$ (see also discussion in [23], p.9). Finally, considering the *preliminary* estimates $\hat{F}_t$, the VAR(p) parameters in (2)

[2]The subscript 0 is used to denote the "true" parameters.

Algorithm 1 Kalman Filter Algorithm

State Equation: $F_t = AF_{t-1} + w_t$ with $w_t \sim N(0, Q), \forall t$
Meas. Equation: $X_t = HF_t + \xi_t$ with $\xi_t \sim N(0, \Psi_d), \forall t$
Initialize:
$F_0 = 0$ and $P_0 = I$
for $t \in \{1, \dots\}$ **do**
  $\{$*Time Updates:*$\}$
  $\hat{F}_{t|t-1} = A\hat{F}_{t-1}$ and $P_{t|t-1} = AP_{t-1}A^\top + Q$
  $\{$*Measurement Updates:*$\}$
  $\Sigma_{t|t-1} = HP_{t|t-1}H^\top + \Psi_d$
  $K_t = P_{t|t-1}H^\top(HP_{t|t-1}H^\top + \Psi_d)^{-1}$
  $\hat{F}_t = \hat{F}_{t|t-1} + K_t(X_t - H\hat{F}_{t|t-1})$
  $P_t = (I - K_tH)P_{t|t-1}$
end for

can be estimated to obtain the matrix polynomial $A(L)$ and the covariance matrix Q.

3) Kalman Filtering: Albeit the misspecification of the "true" model parameters, the theory established in [23] affirms that PCA-based estimation of factors followed by Kalman smoothing leads to *consistent estimation of the state-space* model in (1). Per [23], the Kalman algorithm is employed to *reestimate* the common factors for $t = 1, \dots, T$, given the observations X_t and the dynamics of the factors described via the VAR model of the factors' preliminary estimates (see (2) and (3)). Given the factor model introduced in (1) and (2) (recall that we impose a VAR(p) model on the factors), our state-space representation is described with the *measurement* and *state* equations,

$$X_t = \underbrace{\begin{pmatrix}\hat{\Lambda} & 0 & \cdots & 0\end{pmatrix}}_{H} \begin{pmatrix} F_t \\ F_{t-1} \\ \vdots \\ F_{t-p+1} \end{pmatrix} + \xi_t$$

$$\begin{pmatrix} F_t \\ F_{t-1} \\ \vdots \\ F_{t-p+1} \end{pmatrix} = \underbrace{\begin{pmatrix} A_1 & \cdots & A_{p-1} & A_p \\ I_r & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_r & 0 \end{pmatrix}}_{A} \begin{pmatrix} F_{t-1} \\ F_{t-2} \\ \vdots \\ F_{t-p} \end{pmatrix} + \begin{pmatrix} I_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} w_t.$$

The Kalman smoothing iterative process [46], [47] is illustrated in Algorithm 1. In our proposed system, as new observations X_t, $t = T + 1, \dots$ arrive, the Kalman filter provides a prediction/update $\hat{F}_{t|t-1}$ for the system's "unobserved state", namely the common factors at time t, based on the history of observations up to $t - 1$. Using this "nowcast" for the system state, one can then *obtain a prediction for the power consumption* for time t, i.e., $\hat{X}_t := H\hat{F}_{t|t-1}$. The vector

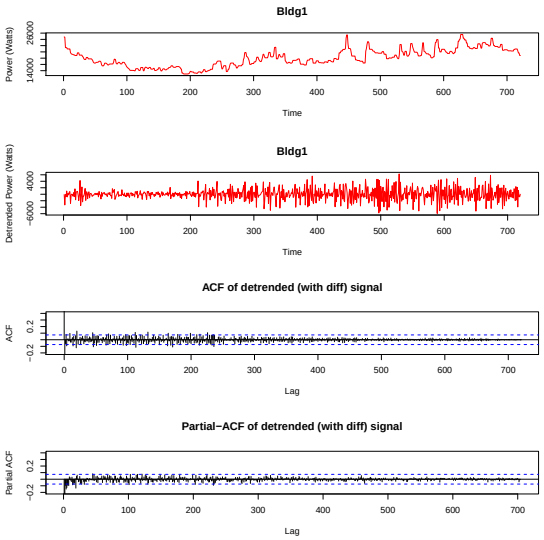

Fig. 5: ACF and partial-ACF of de-trended signal illuminates the suitability of AR(p) modeling. Useful estimates for ACF $\rho(i)$ can be made when $T > 50$ and $i \leq N/4$ [42]. Here, $T = 720$ and $i \leq 50$.

of forecasts for the observed variables is then passed in the hypothesis testing module, described in section III-E.

C. Univariate Autoregressive Models

Next, we present autoregressive models[3] studied. A univariate process (x_t) is said to be an AR(p) process, if x_t is stationary and for every t,

$$x_t - \phi_1 x_{t-1} - \cdots \phi_p x_{t-p} = w_t, \tag{5}$$

where ϕ_i are scalar model coefficients and $w_t \sim WN(0, \sigma^2)$ (WN denotes white noise). We focus our attention to AR(p), rather than a mixed ARMA(p, q) model that would require more parameters/coefficients to be estimated. As illustrated in Fig. 5, the partial-autocorrelation function "tails" off, while the autocorrelation function rapidly vanishes [42]. This was the typical behavior observed for a vast majority of our buildings, and justifies the choice of AR(p) modeling.

Note that the univariate AR model focuses only on the temporal dependence of each time series and does not utilize any information from the other time series under consideration.

For each building, three primary steps were considered: a) *Series transformation:* we applied the first-order differencing operator discussed previously on each

[3]For implementation purposes, we leveraged the `arima` and the `MTS` [43] packages in R for AR/VAR modeling (see our code [48]).

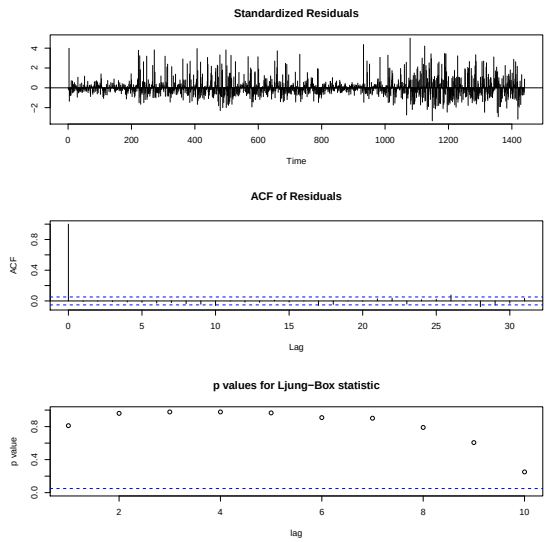

Fig. 6: Model diagnostics for AR(p) models.

building. Variance stabilizing transformations [41] (e.g., square root or logarithm) were also considered but did not offer noticeable improvements; b) *Model selection:* we employ the BIC criterion [41] for model ranking. As depicted in Fig. 7, the maximum lag considered for training AR(p) models was 20; c) *Goodness of fit testing:* for each fitted model, a diagnostics test is performed. We utilize *portmanteau* tests like the Ljung-Box one [41], [42] to assess the fit. Fig. 6 shows a model that passes all diagnostics. Notice that the residuals show no apparent trends and that their ACF function resembles one of white noise. Further, the periodogram in Fig. 4 verifies absence of periodicities. Fig. 7 shows the histogram of the selected AR orders when fitting all meter data with an AR(p) model (using a 2-day training period).

With the estimated coefficients $\hat{\phi}_1, \ldots, \hat{\phi}_p$ from the training period, one can then perform predictions $\hat{x}_t = \hat{\phi}_1 x_{t-1} + \cdots \hat{\phi}_p x_{t-p}$, as new data arrive, in a sequential manner. The prediction error, $e_t := x_t - \hat{x}_t$, is tracked, and its distribution $e_t \sim N(0, \hat{\sigma}^2)$ ($\hat{\sigma}$ is the estimated standard deviation of the training-set residuals) is utilized to compute the monitoring statistic passed to the EWMA control charts discussed in the sequel.

D. Vector Autoregressive Models

VAR(p) models have good properties for (long-term) forecasting[4] of power consumption data and other

[4]We emphasize, though, that our primary focus is rapid outlier detection (false data attacks) rather than longer-term predictions.

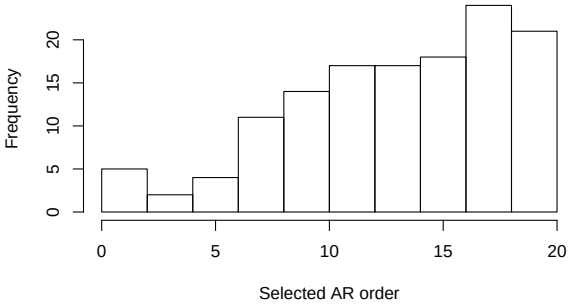

Fig. 7: Histogram of selected order in AR(p) models.

SCADA states [36], [49]. An N-dimensional, zero-mean, stationary process X_t modeled as VAR(p) is given by

$$X_t - \phi_1 X_{t-1} - \cdots \phi_p X_{t-p} = w_t, \tag{6}$$

where ϕ_i are $N \times N$ transition matrices, and (w_t) is a vector white noise process with zero mean and covariance Σ.

Note that a VAR model focuses on learning simultaneously both the temporal dependence of each time series under consideration, as well as cross-temporal dependencies between the time series. Hence, it utilizes lag-information from related time series to improve the forecasting ability for each time series in the model.

To learn the system coefficients, we follow the same training steps as for AR(p) outlined above (i.e., first order differencing, order selection based on BIC, and model checking; we employ a multivariate portmanteau test similar in nature with the Ljung-Box statistic—see Theorem 2.6 [43]). Note that VAR models suffer from the "curse of dimensionality"; there are pN^2 parameters to be learned (for the coefficient matrices ϕ_i, that is). To alleviate the large dimensionality issue we judiciously elect the length for the training period. In particular, we choose T so that $T > pN^2$ (i.e., our data points should be larger than the number of unknown coefficients). Further, rather than working with the full set of buildings, we work with clusters of size $k \ll N$. We allocate buildings in the same cluster based on *Granger causality*, i.e., we group buildings together based on a causality matrix that we compile (see Appendix). Finally, we perform a *model simplification* step (see section 2.7.3 [43]) to refine our trained model to one where parameters that are statistically insignificant are set to zero. For large VAR models, an alternative to the latter step can be performed by fitting a sparse VAR model (for details see [50], [51]).

Once the estimates are obtained, predictions for $t = T+1, \ldots$ are calculated as $\hat{X}_t = \hat{\phi}_1 X_{t-1} + \cdots \hat{\phi}_p X_{t-p}$. The forecasting residuals, $e_t := X_t - \hat{X}_t$, are subsequently used by the hypothesis testing framework, along with the error covariance $\hat{\Sigma}$. All modeling details are available at [48].

E. Sequential Hypothesis Testing

To detect "bad data" attacks, we consider the difference between the forecast provided by the state-space models above and the actual power consumption value reported by the network meters. The distribution of the *error* $e_t := X_t - \hat{X}_t$, under the hypothesis of no anomalies (i.e., the "Null Hypothesis"), can be obtained for all three models.

For illustrative purposes, consider the DFM model (see section III-B). The measurement update step keeps track of the covariance matrix of the measurement error, i.e.,

$$e_t \sim N(0, \Sigma_{t|t-1}). \tag{7}$$

To diagnose anomalies for building j, we focus on e_{jt}, the j-th component of the error. We also consider the variance of the measurement error for that building, updated at each iteration of the Kalman filter, $\Sigma_{t|t-1}(j,j)$. With this information available, we can calculate the p-value of the error obtained. To increase our confidence level when raising an alert, and in an effort to also moderate the false positives rate, we employ an Exponentially Weighted Moving Average (EWMA) control scheme [52], known as *Q-charting* in quality control.

In particular, our methodology considers the normal score $e_{jt}/\Sigma_{t|t-1}(j,j)$ (viz., z-score), and passes the sequence of z-scores in an EWMA control chart methods for detecting "out-of-control" values [52], [53]. Event detection is based on thresholding

$$S_t = (1-\lambda)S_{t-1} + \lambda z_t, \quad \text{where } z_t = e_{jt}/\Sigma_{t|t-1}(j,j)$$

for a weight λ in $(0, 1]$. Both the magnitude and duration of the anomalous event can drive the value of S_t to a level where an alert is triggered. For example, abrupt power shifts would be almost instantaneously detected with high probability. On the other hand, "stealthy" power shifts could be unnoticeable for awhile, but as their duration persists detection probability elevates.

The sensitivity of EWMA is tuned by the weight λ and the threshold parameter L_{ewma}. An alarm is flagged if $S_t > \sigma_\lambda L_{\text{ewma}}$, with $\sigma_\lambda^2 = \lambda/(2-\lambda)$ (see Fig. 1). [52] provides guidelines on calibrating the control chart by choosing appropriate values of λ and L_{ewma} that balance the time between false alarms (average run-length) and the ability to determine whether the process under control has "shifted" to anomalous regimes of certain magnitude. Extensive experimentation suggests that $(\lambda, L_{\text{ewma}})$ pairs $(.09, 3.538), (.29, 3.686), (.53, 3.714), (.84, 3.719)$ are suitable options for our application (see Table I).

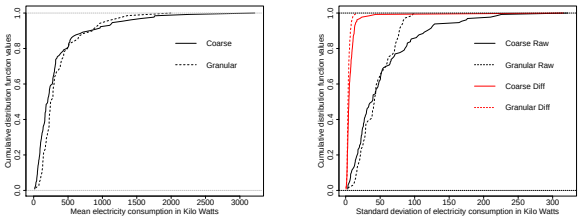

Fig. 8: Cumulative distribution functions for: (L) mean power consumption of buildings, (R) standard deviation of buildings.

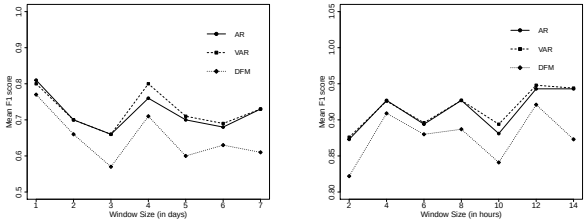

Fig. 9: Sensitivity to training window size.

IV. Performance Evaluation

In this section, we evaluate the proposed framework on synthetic data, as well as real-world smart AMI data obtained from the University of Michigan's (UM) campus infrastructure.

A. Detection accuracy—Synthetic data

To evaluate our framework when the data generating process is known, we *manually* inject false data attacks into synthetically generated data. We simulate different magnitudes, as shown in Table I. For an experiment aiming to detect FDI attacks on building j, we inject, over a period of one hour (we assume 2-min intervals), $M\sigma_j$ Watts, with $M \in \{1.5, 2.5, 3.5\}$. We assess accuracy in terms of the *F1-score*, i.e., the harmonic mean of *precision* and *recall*. Let Tp, Fp and Fn denote the number of *true positives*, *false positives* and *false negatives*, respectively. Precision is defined as

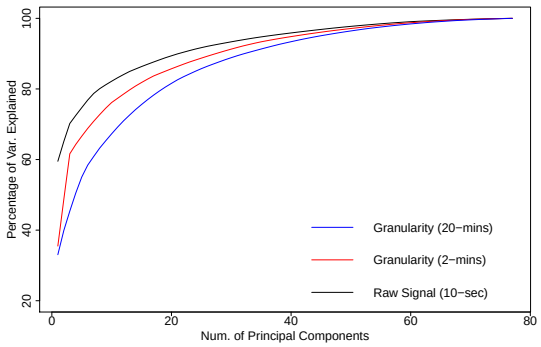

Fig. 10: Percentage of variance captured when various aggregation levels are employed. As shown, more granular datasets exhibit higher correlation among buildings, and few principal components explain a large fraction of the data variability.

$Tp/(Tp + Fp)$ and recall[5] as $Tp/(Tp + Fn)$; both lie in $[0, 1]$.

Table I tabulates the detection performance on synthetic data using the data generation program[6] in [54]. The data is drawn from a DFM with $r = 2$ factors; the factor dynamics are given by a VAR(1) process (see (1) and (2)). We generate data for 130 variables. The results provide important insights: the DFM model (i.e., the model that captures more accurately the data generating process) outperforms all models on detecting low-volume and stealth attacks of small mean shifts. VAR models are not as accurate as DFM models on small magnitudes, but clearly outperform AR models. The latter finding is expected, since the data across buildings exhibit correlations that are used by the multivariate (DFM, VAR) models, which in turn leads to more accurate error detection.

B. Detection accuracy—Real-world data

The real-world dataset consists of power consumption data of about 170 UM buildings. We elaborate on two versions of the data: a) the "coarse granularity" set includes time-series of power consumption at 2-minute intervals; b) the "fine granularity" dataset comprises of time-series at 10-second intervals. After excluding buildings with erratic behavior (e.g, large intervals of missing data, unreasonably high spikes or long stretches of repeated values), the former set contains 130 buildings and the latter 69. Our data includes dorms, office buildings, athletic centers, teaching facilities, etc. For illustration purposes, we utilize data from January 2015.

To visually explore our data, Fig. 3 depicts an image illustrating power load time-series over four days. Further, Fig. 8 provides insights into the average power consumption for our dataset, and the data variability. In addition, Fig. 10 depicts the amount of variance explained, revealed by PCA analysis on the de-trended (mean-adjusted) version of our data. The plots illustrate the "power spectrum" at different aggregation levels. The spectrum reveals that about 20–30 factors explain around 80–90% of the variability in data[7]. This justifies further our claim (see also Fig. 2) that AMI data from a neighborhood-area network are correlated, and provides data-driven justification for the proposed framework and models.

To evaluate our methods with these real-world data, we consider two types of *threat models*: an "agnostic attacker" model in which the malware that compromises a meter has no access to historical meter data, and a "variance-aware" model in which the attacker has access to historical information, and thus may launch seemingly innocuous low-magnitude attacks (e.g., attacks with low signal-to-noise ratio). We train our models using a window of 4 days (12 hours) for the coarse (fine) granularity dataset; these training window sizes were chosen after a comprehensive sensitivity analysis (see Fig. 9).

We elect the number of factors r for the DFM model by inspecting the "screeplot" [43] (i.e., the power spectrum). The screeplot illustrates the amount of variance explained by each principal component. We found that a good cut-off value is $r = 30$ (see Fig. 10). Information theoretic criteria for electing r are discussed in [44].

Tables II–V tabulate detection performance. For each EWMA pair and attack magnitude setting, we report the mean F1-score taken over a series of 200 Monte Carlo experiments. In each experiment, we randomly choose the building under attack and the attack location; we use 720 data points as our "test" dataset. Several interesting observations are in order: a) all three models are robust and detect the injected attacks in the vast majority of scenarios; at the same time they experience a low false positive rate (in fact, alerts are more scarce with the DFM models—see also Fig. 11); b) the VAR model ex-

[5]E.g., assume that a method raised an alert 5 time points after the onset of an attack (recall that we inject one attack that spans 30 data points). In this case, we consider the attack is successfully detected, and set $Tp = 25$, $Fn = 5$. In addition, assume that our algorithm raised 7 false alerts (i.e., $Fp = 7$). We then have $F1 = 0.806$.

[6]See function `synthetic` in our code [48].

[7]The spectrum illustrated here is on *standardized* data. This normalization step is necessary since buildings operate at vary different consumption levels; if not applied, only 1–2 factors would explain most of the variability that would be dominated by the buildings with higher consumption and hence higher variability.

TABLE I: Detection accuracy on synthetically generated data. We report the average F1-score from a series of 100 realizations (standard deviation is shown in parentheses).

Shift ($\times\sigma$)	EWMA	AR	VAR	DFM
1.5	(.09, 3.538)	$.42_{(.32)}$	$.69_{(.17)}$	$\mathbf{.71}_{(.06)}$
2.5	(.09, 3.538)	$\mathbf{.76}_{(.10)}$	$.75_{(.05)}$	$.70_{(.04)}$
3.5	(.09, 3.538)	$\mathbf{.80}_{(.07)}$	$.75_{(.05)}$	$.69_{(.05)}$
1.5	(.29, 3.686)	$.29_{(.36)}$	$.66_{(.29)}$	$\mathbf{.83}_{(.06)}$
2.5	(.29, 3.686)	$.73_{(.27)}$	$\mathbf{.89}_{(.06)}$	$.87_{(.02)}$
3.5	(.29, 3.686)	$\mathbf{.91}_{(.06)}$	$\mathbf{.91}_{(.02)}$	$.86_{(.02)}$
1.5	(.53, 3.714)	$.15_{(.31)}$	$.54_{(.38)}$	$\mathbf{.82}_{(.12)}$
2.5	(.53, 3.714)	$.61_{(.38)}$	$\mathbf{.91}_{(.10)}$	$.89_{(.03)}$
3.5	(.53, 3.714)	$\mathbf{.93}_{(.10)}$	$\mathbf{.93}_{(.04)}$	$.90_{(.01)}$
1.5	(.84, 3.719)	$.09_{(.25)}$	$.28_{(.36)}$	$\mathbf{.55}_{(.33)}$
2.5	(.84, 3.719)	$.48_{(.46)}$	$.78_{(.31)}$	$\mathbf{.91}_{(.04)}$
3.5	(.84, 3.719)	$.91_{(.20)}$	$\mathbf{.96}_{(.02)}$	$.91_{(.02)}$

hibits superior performance than its competitors for low-volume attacks. This is attributed to the fact that prediction strength (and, hence, detection accuracy) elevates when correlated buildings are used in a VAR setting. Recall also that we meticulously construct clusters of buildings based on the *Granger causality* matrix (see Appendix); c) all models demonstrate higher-ranking performance with the 10-second dataset. This reiterates the importance of high-fidelity data for the task at hand. It also motivates the use of highly-adaptive modeling methods that require re-training every few hours in order to more accurately capture the dynamics of power consumption; d) the DFM model depicts the highest performance in the synthetic data scenario (in which the true data generating process was based on a DFM model of 2 factors) and the fine granularity dataset—this emphasizes that the full-strength of such models can be exploited when data variability can be explained by a small number of factors; e) the *majority voting* ensemble is proven robust and often ranks as the most accurate algorithm.

Fig. 11 shows time series of reported alerts (in white) when each building is monitored by the proposed algorithms. The scenario involves an injected attack of 30KWs, at all buildings, starting at time 500. Unquestionably, the attack is easily identified, with very few false positives. In the same figure (lower row panels), we present *visualization charts* that can be easily interpreted by utility operators, and gauge the severity of the situation. The charts show the total alerts over time, across all buildings. Orchestrated attacks on many neighborhood-area buildings can be quickly pinpointed, while at the same time sporadic, false alerts can be dwarfed. As part

TABLE II: Detection performance on real-world AMI data. Coarse (2-min) dataset. Scenario of "agnostic" attacker.

Shift (KWs)	EWMA	AR	VAR	DFM	MJRT
20	(.53, 3.714)	$.55_{(.45)}$	$\mathbf{.58}_{(.45)}$	$.47_{(.43)}$	$.54_{(.46)}$
30	(.53, 3.714)	$.73_{(.37)}$	$\mathbf{.75}_{(.37)}$	$.68_{(.36)}$	$.74_{(.38)}$
40	(.53, 3.714)	$.80_{(.33)}$	$\mathbf{.81}_{(.32)}$	$.72_{(.34)}$	$.79_{(.34)}$
50	(.53, 3.714)	$.88_{(.21)}$	$\mathbf{.89}_{(.21)}$	$.82_{(.21)}$	$\mathbf{.89}_{(.22)}$
20	(.84, 3.719)	$.58_{(.45)}$	$\mathbf{.59}_{(.45)}$	$.52_{(.42)}$	$.57_{(.46)}$
30	(.84, 3.719)	$.78_{(.35)}$	$\mathbf{.79}_{(.34)}$	$.72_{(.33)}$	$\mathbf{.79}_{(.35)}$
40	(.53, 3.714)	$.85_{(.28)}$	$\mathbf{.86}_{(.27)}$	$.76_{(.30)}$	$.85_{(.28)}$
50	(.53, 3.714)	$.89_{(.21)}$	$\mathbf{.90}_{(.20)}$	$.81_{(.21)}$	$.89_{(.23)}$

TABLE III: Detection performance on real-world AMI data. Coarse (2-min) dataset. Scenario of "variance-aware" attacker.

Shift ($\times\sigma W$)	EWMA	AR	VAR	DFM	MJRT
2	(.53, 3.714)	$\mathbf{.19}_{(.35)}$	$.18_{(.35)}$	$.14_{(.30)}$	$.16_{(.34)}$
3	(.53, 3.714)	$.31_{(.43)}$	$\mathbf{.32}_{(.44)}$	$.22_{(.36)}$	$.29_{(.43)}$
4	(.53, 3.714)	$\mathbf{.55}_{(.46)}$	$\mathbf{.55}_{(.46)}$	$.42_{(.43)}$	$.52_{(.46)}$
5	(.53, 3.714)	$\mathbf{.81}_{(.32)}$	$.80_{(.33)}$	$.69_{(.36)}$	$.80_{(.34)}$
2	(.84, 3.719)	$.20_{(.36)}$	$\mathbf{.22}_{(.37)}$	$.18_{(.32)}$	$.19_{(.36)}$
3	(.84, 3.719)	$.34_{(.45)}$	$\mathbf{.37}_{(.46)}$	$.25_{(.38)}$	$.32_{(.44)}$
4	(.53, 3.714)	$.64_{(.44)}$	$\mathbf{.68}_{(.42)}$	$.52_{(.42)}$	$.63_{(.47)}$
5	(.53, 3.714)	$.86_{(.26)}$	$\mathbf{.87}_{(.25)}$	$.79_{(.27)}$	$.86_{(.28)}$

of future work, we plan to investigate methods that help operators calibrate the "system-wide threshold" via a chosen statistical significance level.

V. Conclusions

Monitoring the smart electric grid using AMI measurements is a critical, yet demanding task. It requires careful modeling of the available data underlying dynamics. In this paper, we presented a framework that aims to facilitate modeling of this complex cyber-physical system, and focused on the problem of detecting FDI attacks.

We explored three different statistical models for modeling such data, along with a majority voting scheme, to tackle the problem at hand. Drawing from our experience with correlative patterns among buildings, these models address this issue appropriately. Univariate models are easier to train and implement, and exhibited excellent overall performance. VAR models can capture lead-lag cross-correlations and improve detection accuracy, but may not work well in the presence of hundreds of series / meters. In the latter case, sparse VAR alternatives may prove more appropriate. However, considering small "clusters" of buildings (i.e, k=4–10) can be proven advantageous in detecting anomalies. Further, the dimensionality issue can be addressed with DFM models; they can work well with hundreds of time series, but are more appropriate when strong correlation structures exist (see Table I).

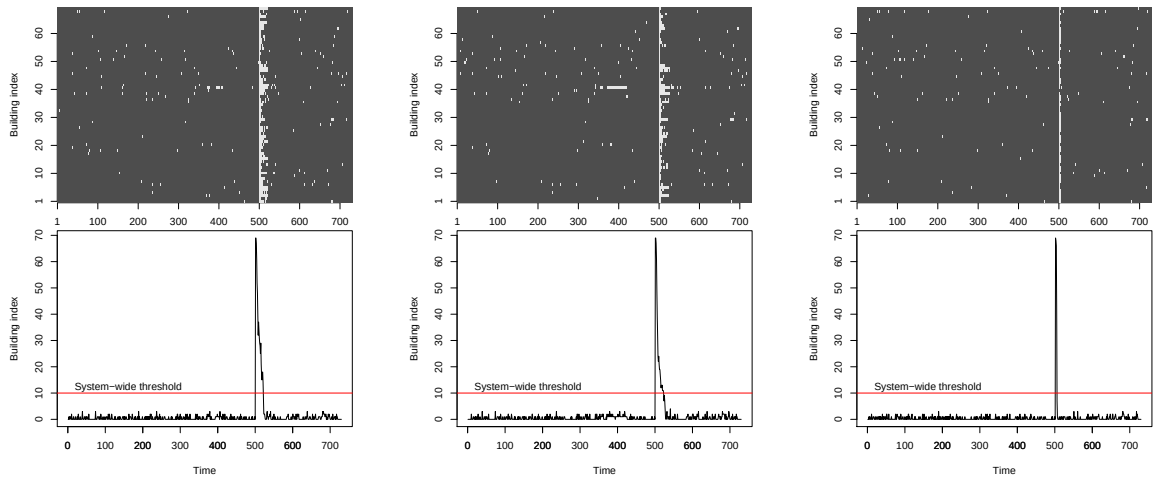

Fig. 11: Visualizations of alerts (granular data, 2 hours). All buildings are monitored simultaneously, and an attack (30KW shift) is injected on all at time 500. Alerts indicated with white. (L) AR model; (M) VAR model; (R) DFM model.

TABLE IV: Detection performance on real-world AMI data. Fine-granularity (10-sec) dataset. "Agnostic" attacker.

Shift (in KWs)	EWMA	AR	VAR	DFM	MJRT
20	(.53, 3.714)	$.86_{(.28)}$	$\mathbf{.87}_{(.28)}$	$.80_{(.33)}$	$.86_{(.29)}$
30	(.53, 3.714)	$.92_{(.14)}$	$\mathbf{.93}_{(.14)}$	$.90_{(.18)}$	$\mathbf{.93}_{(.13)}$
40	(.53, 3.714)	$\mathbf{.93}_{(.13)}$	$\mathbf{.93}_{(.11)}$	$.91_{(.16)}$	$\mathbf{.93}_{(.13)}$
50	(.53, 3.714)	$\mathbf{.94}_{(.07)}$	$\mathbf{.94}_{(.08)}$	$.92_{(.06)}$	$\mathbf{.94}_{(.07)}$
20	(.84, 3.719)	$\mathbf{.88}_{(.24)}$	$\mathbf{.88}_{(.25)}$	$.82_{(.29)}$	$\mathbf{.88}_{(.25)}$
30	(.84, 3.719)	$.92_{(.12)}$	$\mathbf{.93}_{(.12)}$	$.89_{(.17)}$	$\mathbf{.93}_{(.12)}$
40	(.53, 3.714)	$.93_{(.07)}$	$.93_{(.10)}$	$.90_{(.14)}$	$\mathbf{.94}_{(.09)}$
50	(.53, 3.714)	$.93_{(.07)}$	$.93_{(.08)}$	$.91_{(.09)}$	$\mathbf{.94}_{(.06)}$

TABLE V: Detection performance on real-world AMI data. Fine-granularity (10-sec) dataset. "Variance-aware" attacker.

Shift ($\times\sigma W$)	EWMA	AR	VAR	DFM	MJRT
2	(.53, 3.714)	$\mathbf{.31}_{(.42)}$	$.30_{(.41)}$	$.22_{(.37)}$	$.28_{(.41)}$
3	(.53, 3.714)	$.67_{(.42)}$	$\mathbf{.69}_{(.41)}$	$.57_{(.44)}$	$.66_{(.43)}$
4	(.53, 3.714)	$.85_{(.28)}$	$\mathbf{.86}_{(.27)}$	$.77_{(.34)}$	$.85_{(.29)}$
5	(.53, 3.714)	$.93_{(.10)}$	$.93_{(.10)}$	$.90_{(.18)}$	$\mathbf{.94}_{(.10)}$
2	(.84, 3.719)	$.40_{(.44)}$	$\mathbf{.40}_{(.47)}$	$.32_{(.42)}$	$.38_{(.44)}$
3	(.84, 3.719)	$.74_{(.37)}$	$\mathbf{.75}_{(.37)}$	$.64_{(.41)}$	$.73_{(.39)}$
4	(.53, 3.714)	$.88_{(.22)}$	$.88_{(.23)}$	$.84_{(.25)}$	$\mathbf{.89}_{(.22)}$
5	(.53, 3.714)	$.93_{(.07)}$	$.93_{(.09)}$	$.90_{(.14)}$	$\mathbf{.94}_{(.07)}$

Acknowledgments: We thank M. Morgan, Y. Kebede, D. Vorva, E. Cartwright, M. Weiman, N. Remley and O. Webb for their valuable assistance in collecting the data. This work is partially supported by the National Science Foundation CNS-1422078 grant (MK, GM).

APPENDIX

Following [43] (see section 2.2.1), a process x_{2t} is said to *granger cause* another process x_{1t} if predictions for x_{1t} improve significantly in the presence of information on x_{2t}. More formally, denote the available information based on the univariate time series, x_{1t} as $F_1(t) = \{x_{1t}, x_{1t-1}, \cdots\}$. In addition, information for the bivariate vector (x_{1t}, x_{2t}) is represented by $F_2(t) = \{x_{1t}, x_{2t}, x_{1t-1}, x_{2t-1}, \cdots\}$. Let $\hat{x}^{(1)}_{1t+h}$ and $\hat{x}^{(2)}_{1t+h}$ denote the h-step ahead prediction of x_{1t+h} based on the information $F_1(t)$ and $F_2(t)$, respectively. Then, the time series x_{2t} is said to granger cause x_{1t} iff

$$\mathrm{Var}(x_{1t+h} - \hat{x}^{(2)}_{1t+h}) < \mathrm{Var}(x_{1t+h} - \hat{x}^{(1)}_{1t+h}). \quad (8)$$

In our setup, we can represent the joint time series of two buildings by a bivariate model of the form

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \cdots \begin{bmatrix} \phi_{p,11} & \phi_{p,12} \\ \phi_{p,21} & \phi_{p,22} \end{bmatrix} \begin{bmatrix} x_{1t-p} \\ x_{2t-p} \end{bmatrix}$$

plus a noise term.

For a bivariate VAR(p) model, condition (8) is equivalent to the hypothesis testing problem,

$$\begin{aligned} H_0 &: \phi_{i,12} = 0 \text{ for all } i = 1, \cdots, p \\ H_a &: \phi_{i,12} \neq 0 \text{ for some } i = 1, \cdots, p \end{aligned} \quad (9)$$

This hypothesis can be evaluated using a χ^2 Wald test statistic, as described in Section 2.7.3.1 in [43]. Small p-values indicate the presence of granger causality and suggest an improvement in prediction power of x_{1t} in the presence of the time series x_{2t}.

We use the aforementioned methodology in the construction of the Granger matrix G for our set up. The $(i,j)^{\text{th}}$ entry of the matrix G is given by the p-value of the granger causality test based on the bivariate vector $(x_{1t}, x_{2t}) = (X_i(t), X_j(t))$, where $X_i(t)$ and $X_j(t)$ are the univariate electricity consumption time series for the buildings i and j respectively. For a given building i, we choose $k-1$ buildings with smallest entries

corresponding to the i^{th} row of the matrix G. We then use the VAR model based on cluster of these k buildings to predict the electricity consumption for building i.

REFERENCES

[1] U.S. DOE, "Deployment status, department of energy's office of electricity delivery and energy reliability (oe)." [Online]. Available: https://www.smartgrid.gov/recovery_act/deployment_status/

[2] J. John, "US Smart Meter Deployments to Hit 70M in 2016, 90M in 2020." [Online]. Available: goo.gl/ZGTQrR

[3] H. Gharavi and R. Ghafurian, "Smart Grid: The Electric Energy System of the Future," *Proceedings of the IEEE*, June 2011.

[4] Massachusetts Institute of Technology, *The Future of the Electric Grid: An Interdisciplinary MIT study*. MIT Energy Initiative, 2001.

[5] G. B. Giannakis, V. Kekatos, N. Gatsis, S. J. Kim, H. Zhu, and B. F. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 107–128, Sept 2013.

[6] U.S. Department of Energy, Office of Electricity Delivery and Energy Reliability, "Demand Response," www.energy.gov.

[7] A. Ipakchi and F. Albuyeh, "Grid of the future," *Power and Energy Magazine, IEEE*, vol. 7, no. 2, pp. 52–62, 2009.

[8] S. Caron and G. Kesidis, "Incentive-based energy consumption scheduling algorithms for the smart grid," in *IEEE SmartGridComm*, 2010, pp. 391–396.

[9] K. Spees and L. B. Lave, "Demand response and electricity market efficiency," *The Electricity Journal*, vol. 20, no. 3, pp. 69 – 85, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1040619007000188

[10] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *Security Privacy, IEEE*, vol. 7, no. 3, pp. 75–77, 2009.

[11] A. Metke and R. Ekl, "Security technology for smart grid networks," *Smart Grid, IEEE Transactions on*, vol. 1, no. 1, pp. 99–107, 2010.

[12] N. S. T. Bed and U. DOE, "Study of Security Attributes of Smart Grid Systems - Current Cyber Security Issues," April 2009.

[13] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proceedings of CCS '09*, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 21–32.

[14] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, July 2017.

[15] Z. H. Yu and W. L. Chin, "Blind false data injection attack using pca approximation method in smart grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1219–1226, May 2015.

[16] X. Liu, Z. Bao, D. Lu, and Z. Li, "Modeling of local false data injection attacks with reduced network information," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1686–1696, July 2015.

[17] N. Falliere, L. Murch, and E. Chien, "W32.stuxnet dossier," 2011.

[18] R. Lee, M. Assante, and T. Conway, "German steel mill cyber attack," 2014.

[19] ——, "Analysis of the cyber attack on the ukrainian power grid," 2016.

[20] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: Implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317–3318, July 2017.

[21] R. Berthier, W. Sanders, and H. Khurana, "Intrusion detection for advanced metering infrastructures: Requirements and architectural directions," in *IEEE SmartGridComm*, 2010, pp. 350–355.

[22] C. Doz, D. Giannone, and L. Reichlin, "A two-step estimator for large approximate dynamic factor models based on kalman filtering," *Journal of Econometrics*, vol. 164, no. 1, pp. 188 – 205, 2011, annals Issue on Forecasting. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S030440761100039X

[23] D. Giannone, L. Reichlin, and D. Small, "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, vol. 55, no. 4, pp. 665 – 676, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304393208000652

[24] R. B. Litterman, "Forecasting with bayesian vector autoregressionsfive years of experience," *Journal of Business & Economic Statistics*, vol. 4, no. 1, pp. 25–38, 1986.

[25] T. Schon, F. Gustafsson, and P. J. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, July 2005.

[26] F. Cleveland, "Cyber security issues for advanced metering infrastructure," in *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, pp. 1–5.

[27] W. Wang and Z. Lu, "Survey cyber security in the smart grid: Survey and challenges," *Comput. Netw.*, Apr. 2013.

[28] A. Carcano, I. N. Fovino, M. Masera, and A. Trombetta, "State-based network intrusion detection systems for SCADA protocols: A proof of concept," in *Proceedings of CRITIS'09*, 2010, pp. 138–150.

[29] M. G. Kallitsis, G. Michailidis, and S. Tout, "Correlative monitoring for detection of false data injection attacks in smart grids," in *SmartGridComm*, 2015.

[30] M. G. Kallitsis, S. Bhattacharya, S. Stoev, and G. Michailidis, "Adaptive statistical detection of false data injection attacks in smart grids," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2016, pp. 826–830.

[31] X. Liu, Z. Li, and Z. Li, "Optimal protection strategy against false data injection attacks in power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1802–1810, July 2017.

[32] Y. H. et al., "Bad data injection in smart grid: attack and defense mechanisms," *Communications Magazine, IEEE*, Jan. 2013.

[33] S. Bi and Y. J. Zhang, "Graphical methods for defense against false-data injection attacks on power system state estimation," *Smart Grid, IEEE Transactions on*, vol. 5, no. 3, pp. 1216–1227, May 2014.

[34] Yu Wei et al., "An integrated detection system against false data injection attacks in the smart grid," *Security and Communication Networks*, vol. 8, no. 2, 2015.

[35] A. Sanjab and W. Saad, "Smart grid data injection attacks: To defend or not?" in *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2015, pp. 380–385.

[36] J. Zhao, G. Zhang, M. L. Scala, Z. Y. Dong, C. Chen, and J. Wang, "Short-term state forecasting-aided method for detection of smart grid general false data injection attacks," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1580–1590, July 2017.

[37] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 612–621, March 2014.

[38] S. Li, Y. Yilmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2725–2735, Nov 2015.

[39] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.
[40] R. B. B. et al., "Detecting false data injection attacks on DC state estimation," in *SCS Workshop*, 2010.
[41] P. Brockwell and R. Davis, *Time Series: Theory and Methods*. Springer, 1987.
[42] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, *Time Series Analysis: Forecasting and Control, 5th Edition*. Wiley, 2015.
[43] R. S. Tsay, *Multivariate Time Series Analysis: With R and Financial Applications*. Wiley, 2013.
[44] J. Stock and M. Watson, *Dynamic Factor Models*. Oxford: Oxford University Press, 2010. [Online]. Available: http://www.economics.harvard.edu/faculty/stock/files/dfm_oup_4.pdf
[45] S. J.H. and M. Watson, "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, vol. 97, pp. 1167–1179, 2002. [Online]. Available: http://EconPapers.repec.org/RePEc:bes:jnlasa:v:97:y:2002:m:december:p:1167-1179
[46] G. Welch and G. Bishop, "An introduction to the kalman filter," Chapel Hill, NC, USA, Tech. Rep., 1995.
[47] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1986.
[48] M. Kallitsis, S. Bhattacharya, G. Michailidis, "This paper: Extended version and Software in R." [Online]. Available: https://github.com/Merit-Research/wan-fdia
[49] M. Shahidehpour, H. Yamin, and Z. Li, *Short-Term Load Forecasting*. John Wiley & Sons, Inc., 2002, pp. 21–56. [Online]. Available: http://dx.doi.org/10.1002/047122412X.ch2
[50] S. Basu, G. Michailidis *et al.*, "Regularized estimation in sparse high-dimensional time series models," *The Annals of Statistics*, vol. 43, no. 4, pp. 1535–1567, 2015.
[51] S. Basu, A. Shojaie, and G. Michailidis, "Network granger causality with inherent grouping structure," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 417–453, 2015.
[52] J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: Properties and enhancements," *Technometrics*, vol. 32, no. 1, pp. 1–29, Jan. 1990.
[53] D. Lambert and C. Liu, "Adaptive thresholds: Monitoring streams of network counts," *online, J. Am. Stat. Assoc*, pp. 78–89, 2006.
[54] J. Bai and S. Ng, "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions," *Econometrica*, vol. 74, no. 4, pp. 1133–1150, 2006. [Online]. Available: http://www.jstor.org/stable/3805918